



## On Long Memory Origins and Forecast Horizons

Vera-Valdés, J. Eduardo

*Published in:*  
Journal of Forecasting

*DOI (link to publication from Publisher):*  
[10.1002/for.2651](https://doi.org/10.1002/for.2651)

*Publication date:*  
2020

*Document Version*  
Early version, also known as pre-print

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*  
Vera-Valdés, J. E. (2020). On Long Memory Origins and Forecast Horizons. *Journal of Forecasting*, 39(5), 811-826. <https://doi.org/10.1002/for.2651>

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### Take down policy

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

# On Long Memory Origins and Forecast Horizons

J. Eduardo Vera-Valdés\*

*Department of Mathematical Sciences, Aalborg University, and CREATES.*

---

## Abstract

Most long memory forecasting studies assume that the memory is generated by the fractional difference operator. We argue that the most cited theoretical arguments for the presence of long memory do not imply the fractional difference operator, and assess the performance of the autoregressive fractionally integrated moving average (*ARFIMA*) model when forecasting series with long memory generated by nonfractional processes. We find that high-order autoregressive (*AR*) models produce similar or superior forecast performance than *ARFIMA* models at short horizons. Nonetheless, as the forecast horizon increases, the *ARFIMA* models tend to dominate in forecast performance. Hence, *ARFIMA* models are well suited for forecasts of long memory processes regardless of the long memory generating mechanism, particularly for medium and long forecast horizons. Additionally, we analyse the forecasting performance of the heterogeneous autoregressive (*HAR*) model which imposes restrictions on high-order *AR* models. We find that the structure imposed by the *HAR* model produces better long horizon forecasts than *AR* models of the same order, at the price of inferior short horizon forecasts in some cases. Our results have implications for, among others, Climate Econometrics and Financial Econometrics models dealing with long memory series at different forecast horizons. We show in an example that while a short memory autoregressive moving average (*ARMA*) model gives the best performance when forecasting the Realized Variance of the S&P 500 up to a month ahead, the *ARFIMA* model gives the best performance for longer forecast horizons.

*JEL classification:* C53, C22.

*Keywords:* forecasting, *ARFIMA*, long memory, model confidence set, *HAR* model.

---

---

\*E-mail: [eduardo@math.aau.dk](mailto:eduardo@math.aau.dk) Webpage: <https://sites.google.com/view/veravaldes/>  
Address: Skjernvej 4A, 9220 Aalborg Øst, Denmark.

## 1. Introduction

Long memory analysis deals with the notion of series with strong persistence in the sense of long lasting correlations. One of the first works on strong persistence is due to Hurst (1956). He studied the long-term capacity of reservoirs for the Nile and recommended to increase the height of a dam to be built given his observations on cycles of highs at the river. As found by Hurst, failing to account for the presence of long memory can lead to inaccurate forecasts. If the data is best modelled by a long memory process, then the predictions computed with standard models would be too optimistic, in the sense that they would predict a return to normal events faster than what we would observe in reality. A dam built based on a short memory forecast would be more prone to overflow than one built based on a long memory forecast, hence increasing the risk of a catastrophic event. Hurst's work highlights the importance of developing appropriate forecasting tools to deal with the presence of long memory.

In the time series literature, the *ARFIMA* class of models remains to be the most popular given its appeal of bridging the gap between the stationary *ARMA* models, and the nonstationary *ARIMA* model. Moreover, some effort has been directed to assess the performance of the *ARFIMA* type of models when forecasting long memory processes.

Ray (1993) calculates the percentage increase in mean-squared error (*MSE*) from forecasting fractionally integrated (*FI*) series with *AR* models. She argues that the *MSE* may not increase significantly, particularly when we do not know the true long memory parameter. Crato and Ray (1996) compare the forecasting performance of *ARFIMA* models against *ARMA* alternatives and find that *ARFIMA* models are in general outperformed by *ARMA* alternatives for short forecast horizons. Looking at real data, Martens et al. (2009) show that for daily realized volatility for forecast horizons of up to twenty days, it seems to be beneficial to use a flexible high-order *AR* model instead of a parsimonious but stringent fractionally integrated model. On the other hand, Barkoulas and Baum (1997) find improvements in forecasting accuracy when fitting *ARFIMA* models to Eurocurrency returns series, particularly for longer horizons. By allowing for larger data sets of both financial and macro variables, and considering larger forecast horizons, Bhardwaj and Swanson (2006) find that *ARFIMA* processes generally outperform *ARMA* alternatives in terms of forecasting performance. Thus, there does not seem to be a consensus regarding the forecast performance of the *ARFIMA* model.

One thing that most forecasting comparison studies have in common is the underlying assumption that long memory is generated by an *ARFIMA* process. There are two predominant theoretical explanations for the presence of long memory in the time series literature: cross-sectional aggregation of dynamic persistent micro units (Granger, 1980), and that shocks may be of random duration (Parke, 1999). As argued in Section 2, neither of these sources of long memory imply an *ARFIMA* specification. The question addressed in this paper is if an *ARFIMA* specification serves as a good approximation for forecasting purposes when the long memory generating mechanism is different from the *ARFIMA* model.

Moreover, as argued by Baillie et al. (2012), a practitioner’s goals will generally include making forecasts over both short and long horizons. As an example, the surge of Climate Econometrics as a way to address Climate Change relies on the construction of long horizon forecasts while addressing medium term policy goals. Thus, we analyse the forecasting performance of short and long memory models at several forecast horizons. In particular, we extend previous studies to larger forecast horizons relevant to Climate Change analysis.

This paper proceeds as follows. In Section 2, we present the long memory generating processes considered, and show that the most cited theoretical explanations for the presence of long memory do not imply an *ARFIMA* specification. Section 3 describes the design of the Monte Carlo analysis used for the forecasting study. Section 4 presents the results from the forecasting analysis, while Section 5 discusses them in a bias-variance trade-off context. Moreover, Section 6 shows that the insights gained from the Monte Carlo simulations hold on real data. Finally, Section 7 presents the conclusions.

## 2. Long Memory Generating Processes

In this section, we present the selected processes used to generate long memory. All processes considered are long memory in the covariance sense, see Haldrup and Vera-Valdés (2017) for other definitions. In contrast to the alternatives, the covariance sense relates to the rate of decay of the autocorrelations. In this sense, the fitted models try to mimic the rate of decay of the weight that past observations have on future realizations. In this context, the models this information to produce better forecasts; thus, the covariance sense is a sensible definition of long memory for forecasting purposes.

### 2.1. The ARFIMA Model

As a benchmark, we include the *ARFIMA* process due to Granger and Joyeux (1980), and Hosking (1981) in the analysis. The authors extended the *ARMA* model to include fractional dynamics by considering the process

$$\phi(L)(1-L)^d x_t = \theta(L)\epsilon_t, \quad (1)$$

where  $\epsilon_t$  is a white noise process,  $d \in (-1/2, 1/2)$ , and  $\phi(L)$  and  $\theta(L)$  are polynomials in the lag operator with no common roots, all outside the unit circle. The authors used the standard binomial expansion to decompose the fractional difference operator  $(1-L)^d$  in a series with coefficients  $\pi_j = \Gamma(j+d)/(\Gamma(d)\Gamma(j+1))$  for  $j \in \mathbb{N}$ . Using Stirling's approximation, it can be shown that these coefficients decay at a hyperbolic rate, which in turn translates to slowly decaying autocorrelations.

It is well known that *ARFIMA* processes are long memory by all definitions typically considered in the literature, and are relatively easy to estimate by Maximum Likelihood, see Sowell (1992). Thus, the *ARFIMA* model has become the canonical construction for modelling and forecasting long memory in the time series literature; see Beran (1994), and Baillie (1996) for a review.

For the Monte Carlo analysis, we consider *ARFIMA*(1,  $d$ , 0) processes as a way to incorporate both long and short term dynamics.

### 2.2. Cross-Sectional Aggregation

Granger (1980), in line with the work of Robinson (1978) on autoregressive processes with random coefficients, showed that aggregating *AR*(1) processes with coefficients sampled from a Beta distribution can produce long memory. He considered  $N$  series generated as

$$x_{i,t} = \alpha_i x_{i,t-1} + \varepsilon_{i,t} \quad i = 1, 2, \dots, N;$$

where  $\varepsilon_{i,t}$  is a white noise process with  $E[\varepsilon_{i,t}^2] = \sigma_\varepsilon^2 \forall i \in \{1, 2, \dots, N\}, \forall t \in \mathbb{Z}$ . Moreover,  $\alpha_i^2 \sim \mathcal{B}(\alpha; p, q)$  with  $p, q > 1$ , and where  $\mathcal{B}(\alpha; p, q)$  is the Beta distribution with density given by

$$\mathcal{B}(\alpha; p, q) = \frac{1}{B(p, q)} \alpha^{p-1} (1-\alpha)^{q-1} \quad \text{for } \alpha \in (0, 1),$$

with  $B(\cdot, \cdot)$  the Beta function. Furthermore, define the cross-sectional aggregated series as

$$x_t = \frac{1}{\sqrt{N}} \sum_{i=1}^N x_{i,t}.$$

4

Granger showed that as  $N \rightarrow \infty$ , the autocorrelations of  $x_t$  decay at a hyperbolic rate with parameter  $d = 1 - q/2$ ; thus,  $x_t$  has long memory in the covariance sense.

The cross-sectional aggregation result has been extended in different ways, including to allow for general *ARMA* processes, and to other distributions; see Oppenheim and Viano (2004), Linden (1999), and Zaffaroni (2004).

Haldrup and Vera-Valdés (2017) showed that the long memory generated by cross-sectional aggregation does not correspond to the one associated to the *ARFIMA* model. In particular, they showed that although the long memory by cross-sectional aggregation can be removed by fractional differencing, the resulting series does not belong to the class of linear *ARMA* processes. The question addressed in this paper is whether an *ARFIMA* specification remains useful for forecasting purposes.

### 2.3. Error Duration Model

The error duration model was introduced by Parke (1999). He showed that if the series is the result of the sum of shocks of stochastic duration, then it would exhibit long memory in the form of hyperbolic decaying autocorrelations.

Let  $\varepsilon_s$  be a series of *i.i.d.* shocks with mean zero and finite variance  $\sigma^2$ . Assume that the shock  $\varepsilon_i$  has a stochastic duration of  $n_i \geq 0$  time periods, and thus surviving from period  $i$  until period  $i + n_i$ . Let  $p_k$  be the probability that event  $\varepsilon_i$  survives until period  $i + k$ , and take  $g_{i,t}$  to be the indicator function for the event that the error  $\varepsilon_i$  survives until period  $t$ . Furthermore, define  $x_t$  as

$$x_t = \sum_{s=-\infty}^t g_{s,t} \varepsilon_s.$$

Then, if<sup>1</sup>  $p_k \sim k^{-2+2d}$  as  $k \rightarrow \infty$ ,  $x_t$  will have long memory in the covariance sense.

By properly choosing the error survival probabilities, Parke showed that the autocorrelation function will decay at a rate similar to *FI*( $d$ ) processes. However, the resulting series has dichotomous coefficients that do not correspond to the fractional difference operator.

We follow Parke's specification in the Monte Carlo analysis and consider error survival probabilities that mimic those of the *FI*( $d$ ) model.

Table 1 summarizes the long memory generating mechanisms to be analysed.

---

<sup>1</sup>For two series  $a_t, b_t$ , with  $b_t \neq 0 \forall t$ , we write  $a_t \sim b_t$  if  $\lim_{t \rightarrow \infty} a_t/b_t = 1$ .

Table 1: Long Memory Generating Processes

$ARFIMA(p, d, q)$ (DGP 1)	$\phi(L)(1-L)^d x_t = \theta(L)\varepsilon_t$ $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ $(1-L)^d = \sum_{s=0}^{\infty} \frac{\Gamma(s-d)}{\Gamma(-d)\Gamma(s+1)} L^s$
Cross-Sectional Aggregation (DGP 2)	$x_t = \frac{1}{\sqrt{N}} \sum_{i=1}^N x_{i,t}$ $x_{i,t} = \alpha_i x_{i,t-1} + \varepsilon_{i,t}$ $\alpha_i \sim \mathcal{B}(\alpha; p, q); p, q > 1$
Error Duration Model (DGP 3)	$x_t = \sum_{s=-\infty}^t g_{s,t} \varepsilon_s$ $g_{s,s+k} = \begin{cases} 0 & \text{w.p. } 1-p_k \\ 1 & \text{w.p. } p_k \end{cases}$ $p_k = k^{2d-2}$

### 3. Monte Carlo Design

In this section, we describe the Monte Carlo analysis designed to compare the forecasting performance of *ARFIMA* models against *ARMA* and high-order *AR* models on long memory series generated by the processes described in Section 2.

#### 3.1. Forecast Evaluation

We use the Model Confidence Set (*MCS*) approach of Hansen et al. (2011) to assess the forecasting performance of the selected models. From an initial set of models, the methodology allows us to obtain the superior set at a given confidence level. In this sense, the *MCS* is better suited to compare the forecast performance of a large set of competing models.

The *MCS* algorithm proceeds as follows. From a starting set of competing model,  $\mathcal{M}_0$ , we search for the set of superior models at forecast horizon  $h$ ,  $\mathcal{M}^*$ , defined by

$$\mathcal{M}^* = \{i \in \mathcal{M}_0 \mid E(d_{i,j}^h) \leq 0 \quad \forall j \in \mathcal{M}_0\},$$

where  $d_{i,j}^h$  is the loss differential between models  $i$  and  $j$ .

We obtain  $\mathcal{M}^*$  by sequential elimination. For each long memory generating process, we fit all the competing models in the starting set for a sample size  $T$ . The models are indexed by  $i \in \{1, 2, \dots, m\}$ , and the out of sample forecast from model  $i$  is denoted by  $\hat{y}_{T+k}^i, \forall k \in \{1, \dots, h\}$ . We

rank the models according to their expected loss using one of two loss functions: the mean square error ( $MSE$ ),  $L_{SQ}(y_{T+k}, \hat{y}_{T+k}^i) = (y_{T+k} - \hat{y}_{T+k}^i)^2$ , and the mean absolute deviation ( $MAD$ ),  $L_{AD}(y_{T+k}, \hat{y}_{T+k}^i) = |y_{T+k} - \hat{y}_{T+k}^i|$ .

Define the loss differential between models  $i$  and  $j$  by

$$d_{i,j}^k = L_M(y_{T+k}, \hat{y}_{T+k}^i) - L_M(y_{T+k}, \hat{y}_{T+k}^j),$$

for  $M = SQ, AD$ ;  $i, j \in \{1, 2, \dots, m\}$ . We eliminate the worst performing model at each step, and we continue with the process until we can not reject the null hypothesis of equal loss differentials for all models in the set; that is,

$$H_0 : E(d_{i,j}^k) \leq 0 \quad \forall i, j \in \mathcal{M}.$$

The null is tested by using either the range statistic,  $T_R$ , or the semiquadratic statistic,  $T_{SQ}$ , defined by

$$T_R = \max_{i,j \in \mathcal{M}} \frac{|\bar{d}_{i,j}|}{(\widehat{var}(\bar{d}_{i,j}))^{1/2}} \quad T_{SQ} = \sum_{i \neq j} \frac{(\bar{d}_{i,j})^2}{(\widehat{var}(\bar{d}_{i,j}))^{1/2}}.$$

In the Monte Carlo analysis, we present the percentage number of times each model is contained in  $\mathcal{M}^*$  for each forecast horizon.

Additionally, as another measure of forecast performance, we compute both the out of sample root mean square error ( $RMSE$ ), and the out of sample root mean absolute deviation ( $RMAD$ ) given by

$$RMSE_h^i = \left( \frac{1}{h} \sum_{k=1}^h (y_{T+k} - \hat{y}_{T+k}^i)^2 \right)^{1/2} \quad RMAD_h^i = \left( \frac{1}{h} \sum_{k=1}^h |y_{T+k} - \hat{y}_{T+k}^i| \right)^{1/2},$$

where  $h$  and  $\hat{y}_s^i$  are defined as above. We report the mean of both  $RMSE$  and  $RMAD$  across all replications.

Note that the  $MCS$ , and  $RMSE$  or  $RMAD$  evaluation criteria are complementary by construction. The  $MCS$  measure computed in this way, the percentage number of times each model is contained in the set of superior models, tells us about the success rate of the models; that is, how often do we expect each model to perform well. Meanwhile, the  $RMSE$  and  $RMAD$  criteria measure the average performance of each model. We will see in Section 4 how this distinction becomes relevant when selecting a forecasting model.



### 3.2. Model Selection

This section presents the models considered for the forecasting analysis. Table 2 presents the starting set,  $\mathcal{M}_0$ , for the *MCS* approach explained in Section 3.1.

Table 2: Starting Set $\mathcal{M}_0$		
$FI(d)$	$ARMA(1, 1)$	$HAR(3)$
$ARFIMA(1, d, 0)$	$ARMA(2, 1)$	$AR(22)$
$ARFIMA(0, d, 1)$	$ARMA(1, 2)$	$AR(30)$
$ARFIMA(1, d, 1)$	$ARMA(3, 3)$	$AR(50)$
$ARFIMA(2, d, 1)$	$ARMA(4, 4)$	$I(1)$

Model selection was based on two criteria.

As a first criterion, we use the Bayesian Information Criterion (*BIC*) to select the number of lags to include in both the *ARFIMA* and *ARMA* models in an independent Monte Carlo analysis. The validity of the *BIC* for the class of processes with fractional differencing was proven by Beran et al. (1998). The authors show that for this class of processes the penalty term must tend to infinity simultaneously with the sample size; thus, the Akaike Information Criterion is not consistent while the *BIC* is. Note that we made the lag selection exercise independent from the forecasting analysis to avoid the multiple testing problem.

We allow for a maximum of two lags at both components of the *ARFIMA* model, while the maximum was set to four for the *ARMA* model. We use Maximum Likelihood for the estimation of both classes of models with parameter specifications as reported in Appendix A.

Results from the lag selection exercise, presented in Appendix B, show that not many lags are selected for the *ARFIMA* specification for either component. This suggests that the short term component is not that persistent once we control for the long memory behaviour. For the *ARMA* specification, perhaps not surprisingly, more lags are selected due to the fact that we are not controlling for the long memory behaviour by way of the estimation of the fractional memory  $d$ . Nonetheless, the maximum number of lags selected by the *BIC* is two.

As a second criterion, in addition to the preferred models from the lag selection exercise, we follow previous works on long memory forecasting and consider high-order *AR* processes, *AR*(30) and *AR*(50). Moreover, given the success of the *HAR*(3) model of Corsi (2009) on mimicking long memory behaviour, see for instance Andersen et al. (2007) and Chiriac and Voev (2011), we include both the unconstrained *AR*(22), and the *HAR*(3) models. The *HAR*(3) model is a constrained

$AR(22)$  given by

$$x_t = a_0 + a_1 x_{t-1}^{(f)} + a_2 x_{t-1}^{(w)} + a_3 x_{t-1}^{(m)} + \epsilon_t,$$

where  $x_{t-1}^{(f)} = x_{t-1}$ ,  $x_{t-1}^{(w)} = \frac{1}{5} \sum_{i=1}^5 x_{t-i}$  and,  $x_{t-1}^{(m)} = \frac{1}{22} \sum_{i=1}^{22} x_{t-i}$ .

The  $HAR$  specification has been used to model financial data, it reflects the fact that different agents respond to uncertainty at distinct horizons. In this context, the three components of the model seek to capture the daily ( $x_t^{(f)}$ ), weekly ( $x_t^{(w)}$ ), and monthly ( $x_t^{(m)}$ ) levels of uncertainty.

Note that including the  $HAR(3)$  model allows us to extend Corsi's (2009) results in several directions. We make comparisons against a larger set of models, and we use the  $MCS$  approach, which is better suited for comparisons between multiple alternatives. Also, we include larger forecast horizons, and we remove the uncertainty regarding the presence of long memory in the data by comparing the performance of the  $HAR$  model in simulated long memory series, whereas Corsi used real data.

### 3.3. Monte Carlo Design

All models were estimated by Maximum Likelihood ( $MLE$ ) following the work of Baillie et al. (2012) on long memory estimators for forecasting purposes. The authors find the forecasts based on  $MLE$  to be superior than the ones obtained from local Whittle estimators. Moreover, throughout, we use a large sample size of  $T = 1,000$  to reduce the estimation error, and we consider values of the long memory parameter in the stationary range,  $d \in (0, 1/2)$ . Furthermore, given the rise of Climate Econometrics studies keen on producing long horizon forecasts, we consider it relevant to evaluate forecast performances to horizons as far as  $h = 300$ , which correspond to twenty-five years of monthly forecasts.

Table 3 presents the Monte Carlo design for the forecasting analysis.

Table 3: Monte Carlo Design

<ul style="list-style-type: none"> <li>• Generate series of size <math>T + h</math> using the long memory generating processes considered, Section 2, Table 1. The model calibrations are reported in Table 9 in Appendix A.</li> <li>• Fit by Maximum Likelihood the competing models in the starting set <math>\mathcal{M}_0</math>, Table 2, for a sample size <math>T</math>.</li> <li>• Construct forecasts from each model for horizons <math>h \in \{5, 10, 30, 50, 100, 300\}</math>.</li> <li>• Determine the <math>MCS</math> and compute the <math>RMSE</math> and <math>RMAD</math>.</li> <li>• Repeat the steps above <math>R</math> times, the number of replications.</li> </ul>
<ul style="list-style-type: none"> <li>• After the <math>R</math> replications, report the percentage number of times each model is contained in the <math>MCS</math>, and the mean values of <math>RMSE</math> and <math>RMAD</math>, for each forecast horizon.</li> </ul>

## 4. Monte Carlo Results

In this section, we present the results from the Monte Carlo simulations. The parameters for the simulations are presented in Appendix A. Throughout, for reasons of space, we focus on the *MAD* loss function given that it is less sensitive to large misspredictions, see Hansen et al. (2003). Nonetheless, tables using the *MSE* loss function, reported in an Online Appendix, show similar results.

### 4.1. DGP 1: ARFIMA

As a benchmark, we present in Table 4 and Figure 1 the results from the Monte Carlo analysis for an *ARFIMA*(1,  $d$ , 0) process, *DGP* 1, for  $d = 0.3$ .

<i>DGP</i> 1 $d = 0.3$	h=5		10		30		50		100		300	
	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>
<i>FI</i> ( $d$ )	0.937	0.109	0.964	0.118	0.984	0.171	0.988	<b>0.193</b>	0.995	<b>0.259</b>	1.004	<b>0.315</b>
<i>ARFIMA</i> (1, $d$ , 0)	<b>0.933</b>	0.029	0.962	0.028	0.983	0.026	0.988	0.034	<b>0.994</b>	0.035	<b>1.003</b>	0.055
<i>ARFIMA</i> (0, $d$ , 1)	<b>0.933</b>	0.031	<b>0.961</b>	0.030	<b>0.982</b>	0.025	<b>0.987</b>	0.028	<b>0.994</b>	0.039	<b>1.003</b>	0.078
<i>ARFIMA</i> (1, $d$ , 1)	0.935	0.009	0.963	0.011	0.983	0.018	0.988	0.028	<b>0.994</b>	0.025	<b>1.003</b>	0.053
<i>ARFIMA</i> (2, $d$ , 1)	0.935	0.020	0.963	0.019	0.984	0.033	0.989	0.026	0.995	0.033	1.004	0.061
<i>ARMA</i> (1, 1)	0.944	0.130	0.975	0.137	0.995	0.138	0.999	0.128	1.002	0.107	1.007	0.092
<i>ARMA</i> (2, 1)	0.937	0.023	0.966	0.027	0.988	0.013	0.993	0.025	0.999	0.024	1.006	0.045
<i>ARMA</i> (1, 2)	0.938	0.036	0.968	0.032	0.990	0.039	0.995	0.037	1.000	0.034	1.006	0.043
<i>ARMA</i> (3, 3)	0.937	0.032	0.967	0.034	0.988	0.026	0.993	0.019	0.999	0.025	1.007	0.048
<i>ARMA</i> (4, 4)	0.938	0.040	0.967	0.042	0.988	0.035	0.993	0.037	0.999	0.030	1.007	0.063
<i>HAR</i> (3)	0.936	0.039	0.965	0.045	0.986	0.043	0.992	0.049	0.998	0.084	1.006	0.159
<i>AR</i> (22)	0.939	0.075	0.967	0.070	0.988	0.042	0.992	0.038	0.998	0.037	1.005	0.047
<i>AR</i> (30)	0.941	0.073	0.970	0.072	0.989	0.076	0.993	0.067	0.998	0.060	1.005	0.071
<i>AR</i> (50)	0.948	0.143	0.976	0.127	0.994	0.123	0.997	0.111	1.000	0.099	1.006	0.116
<i>I</i> (1)	1.036	<b>0.212</b>	1.076	<b>0.208</b>	1.113	<b>0.193</b>	1.125	0.183	1.151	0.162	1.200	0.126

Table 4: Mean of the *RMAD* and proportion of times the model is in the *MCS* using the *MAD* loss function and the  $T_R$  statistic at a 95% confidence level.

Table 4 shows that *ARFIMA* models are the preferred specification for all forecast horizons measured by the *RMAD* criterion, which is not surprising given that *DGP* 1 is indeed an *ARFIMA* process. Turning to the *MCS* criterion, note that the no-change *I*(1) model gives the best forecast performance for short horizons, while the *FI*( $d$ ) model is the preferred one for medium and large forecast horizons, and its relative performance increases with the forecast horizon.

The results for the  $I(1)$  model are of particular interest. Note that it is the preferred model by the  $MCS$  criterion for short forecast horizons, while it gives the worst performance by the  $RMAD$  criterion for all horizons. As discussed in Section 3.1, these apparent conflicting results can be explained given the complimentary nature of the forecast evaluation measures. Recalling that the  $RMAD$  is a measure across all replications, the results suggest that when the  $I(1)$  is not in the model confidence set, its forecasts perform badly. Nonetheless, its success rate for short forecast horizons could make it a reasonable alternative in some cases, which will probably relate to the distance from the last observation before the forecast to the overall mean. If the last observation is already near the mean, and given that all models will forecast a return to the mean given the stationarity assumption, the no-change model could provide a good forecast alternative, the mean.

Looking at both criteria, we find that high-order  $AR$  and  $ARMA$  models perform quite well when forecasting a true  $ARFIMA$  process for short forecast horizons. In particular, the  $AR(50)$  and  $ARMA(1, 1)$  fall in the superior set of models more than times than  $ARFIMA$  specifications for  $h = 5$ , and  $h = 10$ ; while the increase in the  $RMAD$  criterion relative to  $ARFIMA$  specifications is not too large. A practitioner could in principle construct a weighted measure between the two criteria depending on the problem at hand and choose to use this class of models.

Nonetheless, for medium and large forecast horizons, the  $FI(d)$  model is the one contained in the  $MCS$  the most, and with a  $RMAD$  close to the minimum. In this sense, the  $FI(d)$  appears to be a good overall model for forecasts at medium and large horizons. The superior performance of the  $FI(d)$  model compared to the correct  $ARFIMA(1, d, 0)$  specification may be explained given the small value of the autoregressive coefficient and the estimation error in the long memory parameter. The table suggests that the  $FI(d)$  model seems to capture enough information for forecasting purposes in the long horizon once the short memory component fades away.<sup>2</sup>

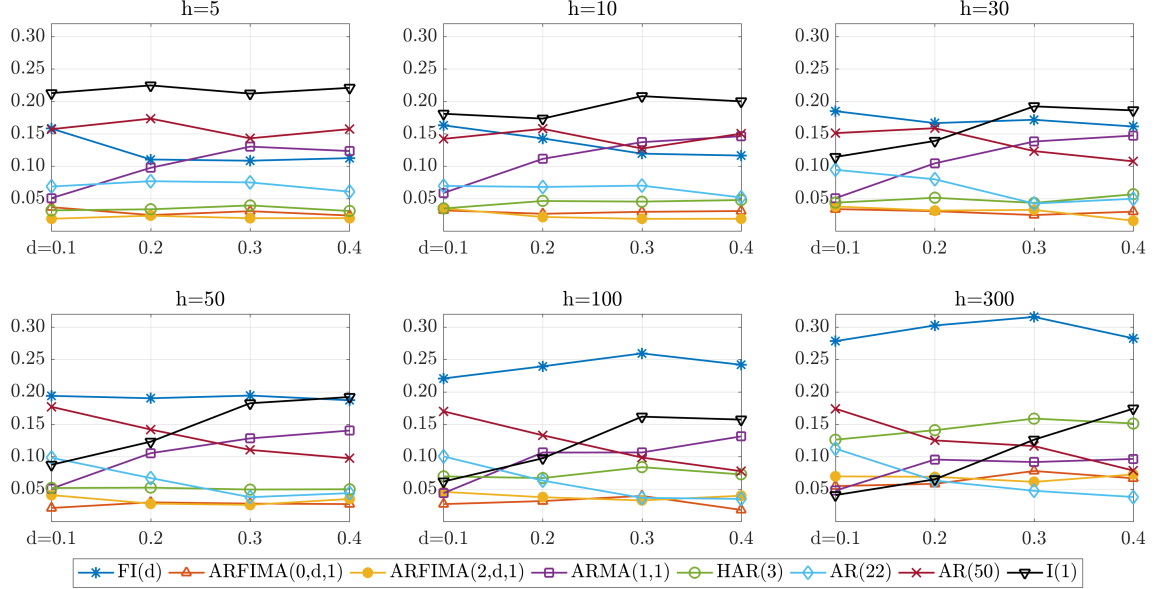
Furthermore, Figure 1 allows us to contrast the performance of high-order  $AR$  models and  $ARFIMA$  models to different degrees of memory.<sup>3</sup> The figure shows that for  $h = 5$  and  $h = 10$ , and for all degrees of memory, the  $AR(50)$  produces better or similar forecast performance according to the  $MCS$  criterion than the  $FI(d)$  model. Yet, the  $FI(d)$  models tend to lead in forecast performance as the horizon increases.

---

<sup>2</sup>Results allowing more short-term dynamics are presented in the Online Appendix.

<sup>3</sup>For ease of exposition, we present a subset of the top performing models in the figures; nonetheless, we present plots with all competing models in the Online Appendix.

Figure 1: Proportion of times the top performing models are in the *MCS* at a 95% confidence level when forecasting *DGP 1* with different degrees of memory at several horizons.



Finally, the figure allows us to compare the *HAR*(3) model against the *AR*(22) model. Note the crossing in preferred model according to the *MCS* criterion between the *AR*(22) and *HAR*(3) models as both the forecast horizon, and degree of memory increase. The figure shows that for  $h = 5$ , the *AR*(22) model is always on top of the *HAR*(3) model. Nonetheless, the preferred model between the two switches from the *AR*(22) to the *HAR*(3) model as the forecast horizon increases. Furthermore, the crossing happens sooner for higher degrees of memory. This suggests that the structure imposed by the *HAR*(3) specification helps to improve forecasting performance for higher degrees of memory, and for larger forecast horizons, at the cost of lower performance at small horizons.

Overall, Table 4, and Figure 1 extend the findings of previous studies on forecasting long memory when the long memory is generated by *ARFIMA* processes. They show that high-order *AR* models are good alternatives for short forecast horizons, while extending the analysis to show that *ARFIMA* models are better suited for medium and large forecast horizons. Moreover, we find that the constraints imposed by the *HAR* model improve forecasting performance over the unconstrained same-order *AR* model for higher degrees of memory, and longer forecast horizons.

#### 4.2. DGP 2: Cross-Sectional Aggregation

Results from the Monte Carlo analysis for the cross-sectional aggregated processes, *DGP 2*, are presented in Table 5, and Figure 2.

<i>DGP 2</i>	h=5		10		30		50		100		300	
$d = 0.3$	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>
<i>FI(d)</i>	1.027	0.172	1.088	0.134	1.161	0.136	1.192	0.135	1.228	0.149	1.267	<b>0.206</b>
<i>ARFIMA(1, d, 0)</i>	<b>1.019</b>	0.036	<b>1.084</b>	0.034	<b>1.159</b>	0.037	<b>1.191</b>	0.037	<b>1.227</b>	0.038	<b>1.266</b>	0.062
<i>ARFIMA(0, d, 1)</i>	1.020	0.046	1.085	0.039	<b>1.159</b>	0.032	<b>1.191</b>	0.047	<b>1.227</b>	0.059	<b>1.266</b>	0.105
<i>ARFIMA(1, d, 1)</i>	<b>1.019</b>	0.020	<b>1.084</b>	0.028	1.161	0.048	1.192	0.045	<b>1.227</b>	0.041	<b>1.266</b>	0.072
<i>ARFIMA(2, d, 1)</i>	<b>1.019</b>	0.007	1.086	0.019	1.164	0.024	1.196	0.040	1.230	0.060	1.268	0.071
<i>ARMA(1, 1)</i>	1.029	0.095	1.097	0.121	1.184	0.142	1.216	0.132	1.243	0.107	1.275	0.084
<i>ARMA(2, 1)</i>	1.022	0.032	1.089	0.034	1.172	0.026	1.205	0.016	1.235	0.028	1.272	0.049
<i>ARMA(1, 2)</i>	1.026	0.024	1.093	0.026	1.178	0.036	1.212	0.035	1.241	0.033	1.274	0.037
<i>ARMA(3, 3)</i>	1.026	0.024	1.092	0.028	1.173	0.022	1.205	0.026	1.237	0.030	1.275	0.054
<i>ARMA(4, 4)</i>	1.024	0.025	1.090	0.028	1.171	0.023	1.204	0.023	1.235	0.029	1.273	0.060
<i>HAR(3)</i>	1.021	0.017	1.087	0.019	1.168	0.052	1.203	0.059	1.236	0.077	1.275	0.173
<i>AR(22)</i>	1.023	0.060	1.089	0.050	1.168	0.033	1.201	0.038	1.233	0.033	1.271	0.040
<i>AR(30)</i>	1.025	0.070	1.092	0.063	1.171	0.047	1.203	0.042	1.234	0.037	1.271	0.035
<i>AR(50)</i>	1.032	0.145	1.097	0.142	1.177	0.120	1.208	0.105	1.237	0.112	1.273	0.071
<i>I(1)</i>	1.075	<b>0.227</b>	1.164	<b>0.235</b>	1.275	<b>0.222</b>	1.326	<b>0.220</b>	1.387	<b>0.208</b>	1.479	0.191

Table 5: Mean of the *RMAD* and proportion of times the model is in the *MCS* using the *MAD* loss function and the  $T_R$  statistic at a 95% confidence level.

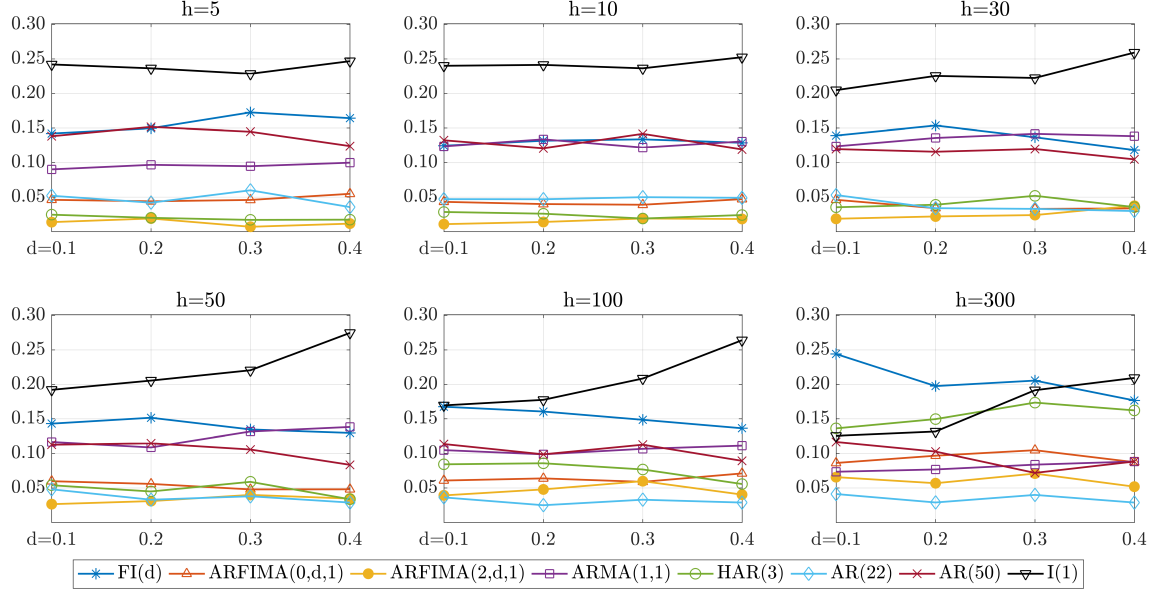
Note that the  $I(1)$  model is the one contained in the superior set of models the most for forecasts horizons up to 100, while performing last according to the *RMAD* criterion. Once again, this seems to suggest a higher success rate for the  $I(1)$  model at short and medium forecast horizons, but at the price of high variability on its performance, which may relate to the distance from the last observed value to the mean.

Looking for a better balance between the criteria, note that the *ARFIMA* class of models is the preferred one according to the *RMAD* criterion for all forecast horizons, with the  $FI(d)$  in particular remaining among the top performing according to the *MCS* criterion. In this sense, a weighted average between the criteria would presumably result in selecting the  $FI(d)$  model as a well suited model for forecasting purposes, specially for medium and large forecast horizons. Among the short memory models, the  $ARMA(1, 1)$  and  $AR(50)$  models could provide good forecast alternatives for short forecast horizons, as seen by their relatively high value for the *MCS* criterion, and *RMAD* relatively close to the minimum.

We can see the effect that the degree of long memory has on the results in Figure 2. We plot the percentage of number of times the models are contained in the *MCS* for different degrees of memory, for all forecast horizons.

The figure extends the findings in Table 5, it shows the good performance of the  $ARMA(1, 1)$

Figure 2: Proportion of times the top performing models are in the *MCS* at a 95% confidence level when forecasting *DGP 2* with different degrees of memory at several horizons.



and  $AR(50)$  models at short and medium forecast horizons, providing similar results to the  $FI(d)$  specification. Nonetheless, the figure shows the increase in relative forecast performance of the  $FI(d)$  model when the forecast horizon increases. Also, the plot shows the increase in forecast performance of the  $HAR(3)$  model for large forecast horizons. In particular, while the performance in small forecast horizons is inferior in comparison to the unconstrained  $AR(22)$ , the constraints seem to introduce the additional structure needed for good medium and large horizon forecasts. Section 5 will analyse this feature further.

Overall, Table 5 and Figure 2 indicate that the *ARFIMA* class of models are a good specification for forecast construction when working with long memory series generated by cross-sectional aggregation, *DGP 2*, particularly for medium and large forecast horizons. This in the sense that they obtain a good balance performance among both evaluation criteria. Among the short memory models, the  $AR(50)$  and  $ARMA(1,1)$  could provide sensible alternatives for smaller forecasting periods. Finally, the  $HAR(3)$  model starts to show good performance at larger horizons, with slightly inferior performance at short forecast horizons.

#### 4.3. *DGP 3: Error Duration Model*

Table 6 presents the results from the Monte Carlo analysis for processes generated using the error duration model, *DGP 3*, for long memory parameter  $d = 0.3$ .

<i>DGP 3</i> <i>d = 0.3</i>	h=5		10		30		50		100		300	
	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>
<i>FI(d)</i>	1.097	0.130	1.123	0.126	1.131	<b>0.162</b>	1.133	<b>0.193</b>	1.136	<b>0.217</b>	1.140	<b>0.277</b>
<i>ARFIMA</i> (1, <i>d</i> , 0)	1.075	0.038	1.105	0.037	1.120	0.041	1.124	0.035	1.130	0.042	1.136	0.063
<i>ARFIMA</i> (0, <i>d</i> , 1)	1.074	0.028	1.109	0.028	1.130	0.027	1.137	0.023	1.145	0.029	1.156	0.059
<i>ARFIMA</i> (1, <i>d</i> , 1)	<b>1.067</b>	0.012	<b>1.100</b>	0.012	<b>1.118</b>	0.009	<b>1.123</b>	0.021	<b>1.129</b>	0.021	<b>1.135</b>	0.037
<i>ARFIMA</i> (2, <i>d</i> , 1)	<b>1.067</b>	0.012	<b>1.100</b>	0.016	<b>1.118</b>	0.025	<b>1.123</b>	0.021	<b>1.129</b>	0.031	<b>1.135</b>	0.041
<i>ARMA</i> (1, 1)	1.077	0.129	1.109	0.110	1.125	0.089	1.129	0.082	1.133	0.073	1.137	0.071
<i>ARMA</i> (2, 1)	1.074	0.031	1.108	0.040	1.125	0.043	1.128	0.042	1.133	0.043	1.137	0.050
<i>ARMA</i> (1, 2)	1.070	0.018	1.103	0.020	1.120	0.024	1.125	0.030	1.130	0.030	1.136	0.044
<i>ARMA</i> (3, 3)	1.069	0.023	1.102	0.025	1.119	0.033	1.124	0.033	1.130	0.040	1.136	0.065
<i>ARMA</i> (4, 4)	1.070	0.043	1.103	0.033	1.119	0.039	1.124	0.050	1.130	0.047	<b>1.135</b>	0.079
<i>HAR</i> (3)	1.076	0.100	1.106	0.104	1.121	0.086	1.126	0.095	1.131	0.099	1.137	0.145
<i>AR</i> (22)	1.077	0.052	1.108	0.062	1.122	0.074	1.126	0.066	1.131	0.057	1.136	0.065
<i>AR</i> (30)	1.078	0.074	1.109	0.077	1.124	0.094	1.126	0.089	1.131	0.085	1.136	0.096
<i>AR</i> (50)	1.085	0.146	1.116	0.149	1.131	0.138	1.133	0.125	1.134	0.134	1.137	0.143
<i>I</i> (1)	1.212	<b>0.164</b>	1.261	<b>0.161</b>	1.298	0.116	1.309	0.095	1.324	0.089	1.371	0.068

Table 6: Mean of the *RMAD* and proportion of times the model is in the *MCS* using the *MAD* loss function and the  $T_R$  statistic at a 95% confidence level.

We can see from the table that the *ARFIMA* class of models provide the best performance measured by the *RMAD* criterion for all forecast horizons, while showing the best performance by the *MCS* criterion for  $h = 30$  and larger. Thus, the *ARFIMA* class of models seem to provide the best forecast performance for series generated by the error duration model, *DGP 3*.

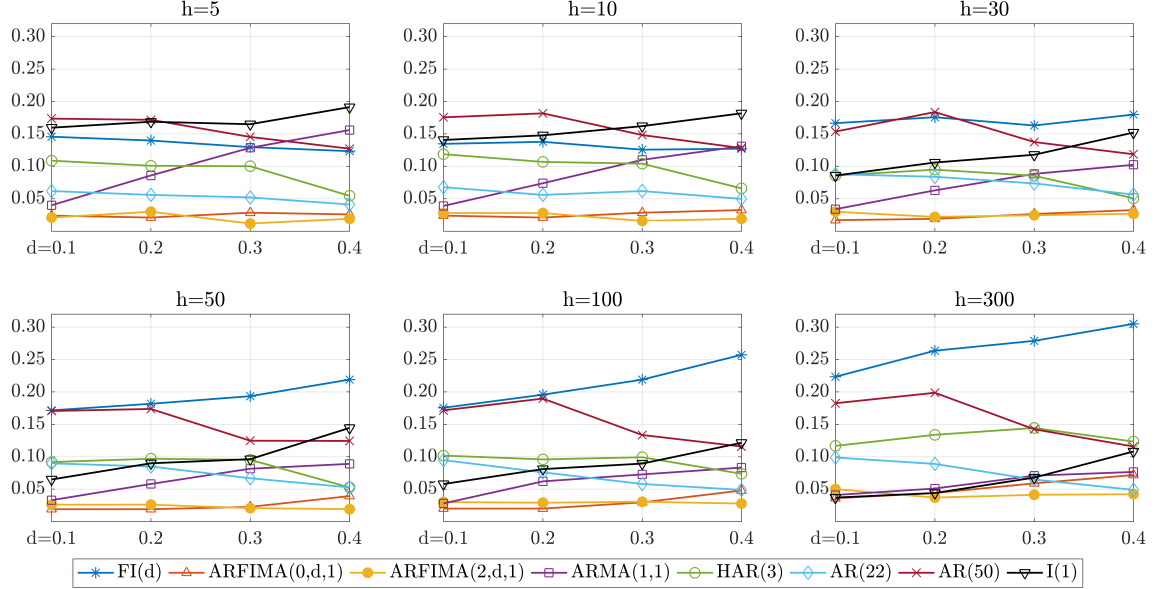
Turning to short memory alternatives, note the relatively good performance of the *AR*(50) model for all forecast horizons. Thus, even though the *FI(d)* model appears as the best performing model, high-order *AR* models seem to produce good short and medium horizon forecast alternatives for *DGP 3*. Moreover, contrasting the performance of the *HAR*(3) model against the *AR*(22) model, the table shows the gains in performance of imposing some structure into the higher-order *AR* models when forecasting *DGP 3*, with similar or better performance on both criteria.

Figure 3 presents the proportion of times the models are contained in the *MCS* when forecasting *DGP 3* for different degrees of memory. The figure shows the relative performance increase of the *FI(d)* model over high-order *AR* models as both the degree of memory, and the forecast horizon increase; thus, it extends the insights gained from the table to all degrees of memory considered. Also, note that the *HAR*(3) model is always on top of the unconstrained *AR*(22) model, if slightly for medium horizons.

Overall, Table 6 and Figure 3 convey one of the main insights from this paper; that is, short memory models are good alternatives for short horizon forecasts, while long memory models are preferred for medium and long horizon forecasts of long memory processes, regardless of the memory generating mechanism.



Figure 3: Proportion of times the top performing models are in the *MCS* at a 95% confidence level when forecasting *DGP 3* with different degrees of memory at several horizons.



## 5. Discussion

Looking at the relative performance between an unconstrained high-order *AR* model and a constrained *HAR*, the results from the Monte Carlo simulation can be further analysed in the context of the bias-variance trade-off typically studied in regression analysis.

All processes considered in this paper are long memory in the covariance sense; hence, the models are fitted to capture the information contained in the autocorrelation function and use it for forecasting purposes. In other words, the models select  $\{a_i\}_{i=0}^T$  in the  $x_t = a_0 + \sum_{i=1}^T a_i x_{t-i}$  representation, with the aim of replicating the autocorrelation function.

*ARFIMA* and *ARMA* models differ in terms of the way to select the coefficients  $a_i$ . *ARFIMA* models impose a hyperbolic rate by the fractional differencing operator  $(1 - L)^d$ , see Equation 1, while high-order *AR* models are more flexible by selecting each coefficient individually. In this sense, the fractional models need just one parameter to establish the infinite list of coefficients, and are hence of low variance. Nonetheless, the uncertainty surrounding the estimation of the long memory parameter may introduce some bias. As an alternative, high-order *AR* models are more flexible. Hence, they can reduce the bias of the coefficients, but suffer from increased variance given the number of estimated parameters. This distinction can be particularly important in the scenario of

having small estimation samples, something we abstract from in this study. Given the uncertainty associated to estimating an increasing number of parameters, we would expect the performance of high-order *AR* models to deteriorate in short series. Yet, as the Monte Carlo analysis showed, this flexibility can produce good forecast performance at short horizons, particularly when the degree of memory is small. Nonetheless, *AR* models lose forecasting power as the forecast horizon gets larger. We could increase the order of the autoregressive process to increase the forecasting performance at long horizons, but the estimation becomes unstable.

In this context, *HAR* models are a compromise between the rigid *ARFIMA* and flexible high-order *AR* model specifications. They incorporate high-order autoregressive specifications while greatly restricting the number of parameters to be estimated. This arrangement allows the *HAR* model to provide similar forecast performance at medium forecast horizons as same-order unrestricted *AR* models, while providing better long horizon forecasts. Yet, as shown in the Monte Carlo analysis, *HAR* models may suffer a forecast performance loss at short horizons.

To further illustrate this point, Figure 4 compares the forecasting performance of constrained *AR* models in the spirit of the *HAR* model against their unconstrained specifications. In particular, in addition to the *HAR*(3) and *AR*(22) models, we show an unrestricted *AR*(50) model and a *HAR*(4) given by

$$x_t = a_0 + a_1 x_{t-1}^{(f)} + a_2 x_{t-1}^{(w)} + a_3 x_{t-1}^{(m)} + a_4 x_{t-1}^{(b)} + \epsilon_t,$$

where  $x_{t-1}^{(f)} = x_{t-1}$ ,  $x_{t-1}^{(w)} = \frac{1}{5} \sum_{i=1}^5 x_{t-i}$ ,  $x_{t-1}^{(m)} = \frac{1}{22} \sum_{i=1}^{22} x_{t-i}$ , and  $x_{t-1}^{(b)} = \frac{1}{50} \sum_{i=1}^{50} x_{t-i}$ . Note that the *HAR*(4) model just described is a constrained *AR*(50).

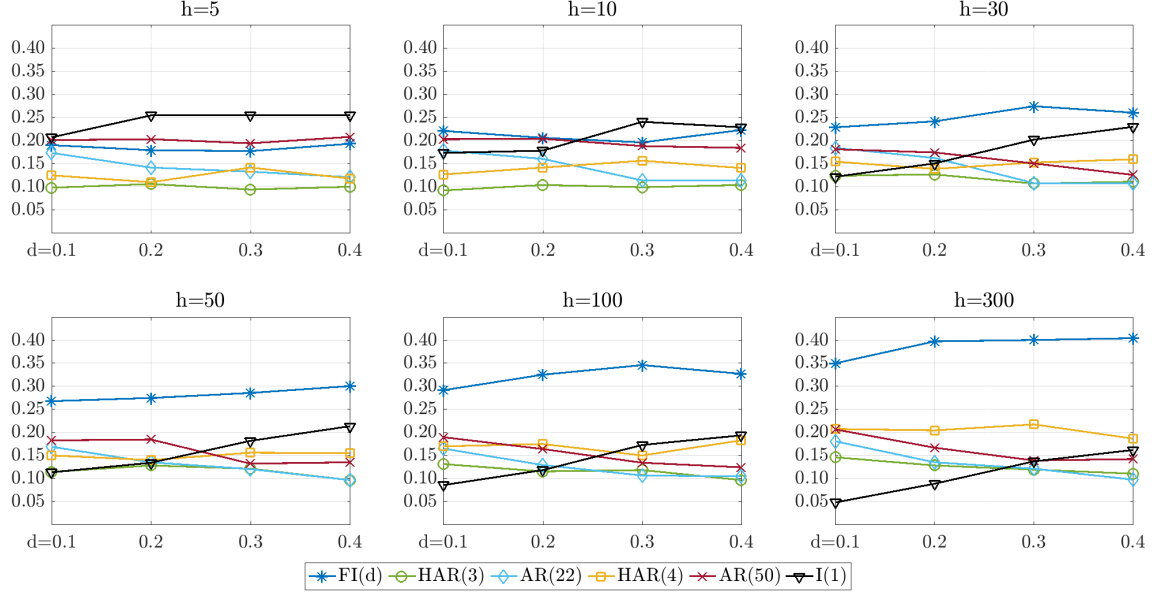
Figure 4 shows the average number of times two high-order *AR* processes and their comparable *HAR* specifications are contained in the *MCS* when forecasting *DGP* 1. The figure displays the superior performance of unconstrained autoregressive processes against constrained ones of the same order for short forecast horizons. Nonetheless, it shows the increase in relative performance for the constrained versions at large forecast horizons. In particular, for  $h = 5$  and  $h = 10$ , the unconstrained *AR* models give better performance than equivalent order *HAR* alternatives for all degrees of memory; while at  $h = 300$ , the *HAR*4 is always on top of the unconstrained *AR*(50).<sup>4</sup>

The bias-variance trade-off has been a topic of great interest in the literature of regressions with a high number of covariates, it thus would be compelling to adapt shrinkage and sparse methods to

---

<sup>4</sup>The Online Appendix shows that this result extends to the other *DGPs* considered.

Figure 4: Proportion of times the models are in the *MCS* at a 95% confidence level when forecasting *DGP 1* with different degrees of memory at several horizons. For the plots, the starting set contains only the six models shown.



lag selection in the context of long memory forecasting. This line of inquiry is left open for future research.

## 6. Illustrative Example

In this section, we evaluate the forecasting performance of the competing models on real data. We select the Realized Variance (*RV*) of the S&P 500 for illustration. As it is well known, the Realized Variance is a measure of volatility. We obtain the *RV* series from the Oxford-Man Institute’s “Realised Library” computed on the basis of intradaily observations spaced into 5-minute intervals and subsampled at a 1-minute frequency. The sample runs from January 3, 2000 until December 30, 2015.

The *RV* has been proven to have long memory by, among others, Martens et al. (2009) and Andersen et al. (2003). In particular, notice that by construction the S&P 500 series is an aggregated measure; thus, it is in line with the cross-sectional argument for long memory. Furthermore, Parke (1999) argues that given the difference between information quality among agents, the error duration model is capable of explaining the long memory in volatility. Hence, realized variance can be argued to have long memory by the theoretical explanations considered in this work, making it a good fit for the exercise.

Figure 5: Realized Variance and its autocorrelation function.

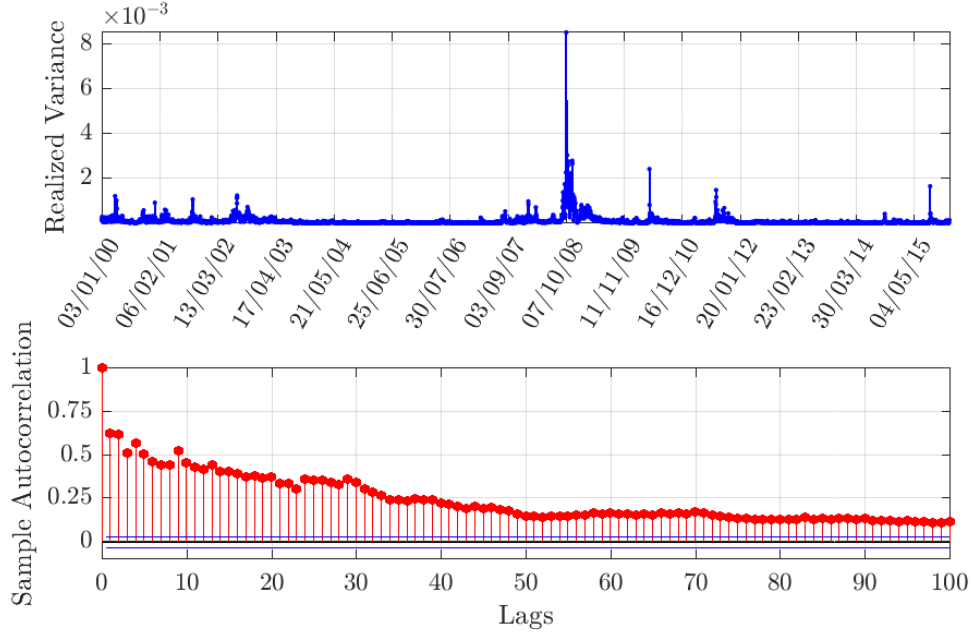


Figure 5 presents the  $RV$  series and its autocorrelation function. The autocorrelation function shows behaviour similar to those of long memory processes, remaining significant at large lags. The estimates for the long memory parameter are 0.4501, 0.4675, and 0.3756 by the semi-parametric estimator of Geweke and Porter-Hudak (1983), the local Whittle estimator of Robinson (1995) and Künsch (1986), and the Maximum Likelihood Estimator of Sowell (1992), respectively.

We use as estimation sample an increasing window starting from the period between January 3, 2000 to December 30, 2014. The last estimation window runs from January 3, 2000 to December 30, 2015. For each estimation window we construct forecasts for 5, 10, 22, 66, 120 and 254 periods ahead, and compute the Model Confidence Set for each estimation window and forecast horizon.

Table 7 presents the results from the forecasting exercise on the first estimation window. As the table shows,<sup>5</sup> short memory models give the best forecast performance for horizons up to 22 periods ahead, in line with the results of Martens et al. (2009). In particular, for short forecast horizons either the  $ARMA(1, 2)$  or the  $ARMA(1, 2)$  is the one with minimum  $RMAD$ , and the only one contained in the  $MCS$ . Nonetheless, the  $ARFIMA$  type of models tends to dominate in forecast performance as the forecast horizon increases, in line with our results from the Monte Carlo

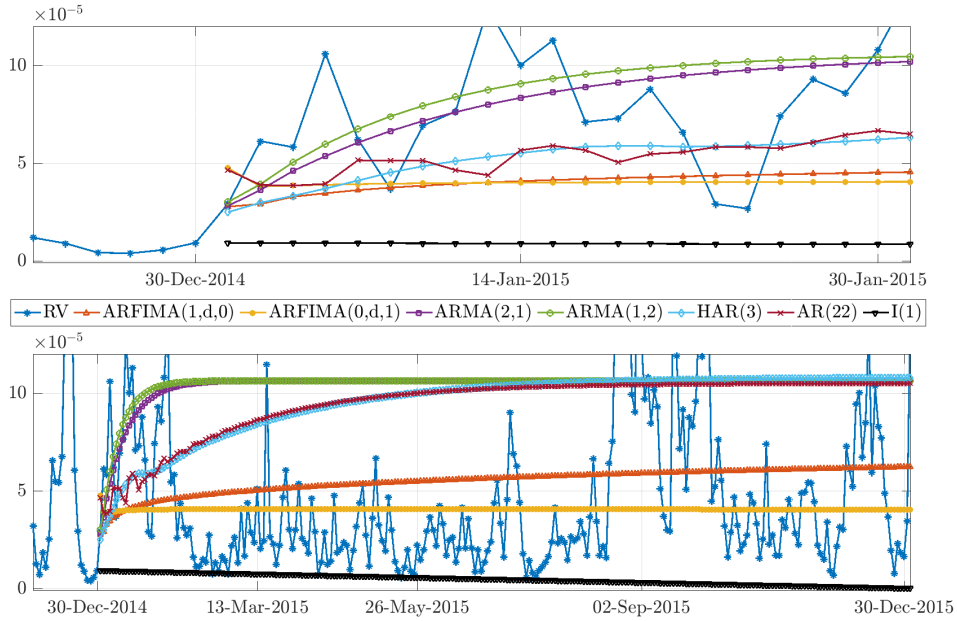
<sup>5</sup>The  $ARFIMA(2, d, 1)$  showed convergence problems and thus was excluded from this exercise.

analysis. For horizons  $h \in \{66, 120, 254\}$ , representing 3, 6 and 12 months, the  $ARFIMA(0, d, 1)$  is the only one contained in the  $MCS$  and with lowest  $RMAD$ .

Realized Variance	h=5		10		22		66		120		254	
	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>	<i>RMAD</i>	<i>MCS</i>
$FI(d)$	0.0049	0	0.0056	0	0.0060	0	<b>0.0054</b>	0	0.0053	0	0.0063	0
$ARFIMA(1, d, 0)$	0.0051	0	0.0057	0	0.0060	0	0.0056	0	0.0057	0	0.0065	0
$ARFIMA(0, d, 1)$	0.0054	0	0.0060	0	0.0063	0	<b>0.0054</b>	<b>1</b>	<b>0.0052</b>	<b>1</b>	<b>0.0062</b>	<b>1</b>
$ARFIMA(1, d, 1)$	0.0054	0	0.0060	0	0.0063	0	<b>0.0054</b>	0	0.0053	0	0.0063	0
$ARMA(1, 1)$	0.0057	0	0.0064	0	0.0065	0	0.0060	0	0.0066	0	0.0076	0
$ARMA(2, 1)$	0.0043	0	0.0044	0	<b>0.0048</b>	<b>1</b>	0.0077	0	0.0083	0	0.0086	0
$ARMA(1, 2)$	<b>0.0041</b>	<b>1</b>	<b>0.0043</b>	<b>1</b>	0.0050	0	0.0078	0	0.0084	0	0.0086	0
$ARMA(3, 3)$	0.0064	0	0.0065	0	0.0060	0	0.0074	0	0.0081	0	0.0085	0
$ARMA(4, 4)$	0.0056	0	0.0058	0	0.0053	0	0.0075	0	0.0082	0	0.0085	0
$HAR(3)$	0.0055	0	0.0057	0	0.0056	0	0.0068	0	0.0077	0	0.0083	0
$AR(22)$	0.0052	0	0.0057	0	0.0057	0	0.0069	0	0.0077	0	0.0083	0
$AR(30)$	0.0057	0	0.0064	0	0.0063	0	0.0064	0	0.0071	0	0.0079	0
$AR(50)$	0.0056	0	0.0062	0	0.0059	0	0.0070	0	0.0079	0	0.0084	0
$I(1)$	0.0074	0	0.0080	0	0.0083	0	0.0061	0	0.0053	0	0.0070	0

Table 7:  $RMAD$  and indicator function for the model being contained in the  $MCS$  using the  $MAD$  loss function and the  $T_R$  statistic at a 95% confidence level.

Figure 6: Realized Variance and forecasts.



The results from Table 7 can be better understood by looking at Figure 6. The figure shows the true  $RV$  and the forecasts from the best competing models. The top plot shows the forecasts up to a month ahead, while the bottom one extends it to a year ahead.

As the figure shows, all models determine that the  $RV$  is away from its mean by the end of

the estimation period. The main difference between models being the speed of convergence to the mean. The fast convergence of the *ARMA* specifications allows them to capture the quick bounce at the beginning of the forecasting period much better than the *ARFIMA* alternatives, which results in better short horizon forecasts. Nonetheless, the long memory alternatives are better suited to capture the long horizon dynamics of the *RV* as the forecast horizon increases.

For robustness, we extend the analysis to the exercise with an increasing estimation window. Table 8 presents the average number of times the competing models are in the superior set of models. For ease of exposition, we have pooled the results between long memory alternatives (*ARFIMA* models) and short memory ones (*ARMA* models).

Realized Variance	$h = 5$	10	22	66	120	254
<i>ARFIMA</i> alternatives	0.268	0.342	<b>0.342</b>	<b>0.425</b>	<b>0.720</b>	<b>0.705</b>
<i>ARMA</i> alternatives	<b>0.409</b>	<b>0.370</b>	0.335	0.264	0.008	0
<i>I</i> (1) no-change model	0.323	0.287	0.323	0.311	0.311	0.335

Table 8: Average number of times the type of model is in the *MCS* at a 95% confidence level for the increasing estimation window exercise. In total, 254 estimation windows were considered.

The table confirms our simulations and first estimation window results. It shows that for the short horizon, the *ARMA* models are the ones contained the most in the *MCS*. As we increase the forecasting horizon, the *ARFIMA* alternatives are the ones contained the most in the *MCS*. In particular, no short memory alternative is contained in the *MCS* for forecast up to a year ahead, for all estimation windows.

## 7. Conclusions

This paper argues that the most cited theoretical arguments behind the presence of long memory in the data do not correspond to the fractional difference operator. In this context, it evaluates the forecasting performance of *ARFIMA* models when the memory is generated from nonfractional sources.

We find that high-order *AR* models produce comparable forecasts as *ARFIMA* models at short horizons. Nonetheless, as the forecast horizon increases, the *ARFIMA* models tend to dominate in terms of forecast performance. Hence, *ARFIMA* models are well suited for medium and long horizon forecasts of long memory regardless of the generating mechanism, while high-order *AR* models may be good alternatives for forecasts at short horizons. In particular, we find that if the

long memory is generated by the error duration model, the fractionally integrated model produces the best forecast performance at medium and large horizons for all degrees of memory, while remaining competitive at short horizons.

Additionally, by making a compromise between flexibility and complexity, we find that the structure imposed by the *HAR* model induces a trade-off in forecast performance at different forecast horizons. In other words, the *HAR* model produces better long horizon forecasts, similar medium horizon forecasts, and similar or inferior short horizon forecasts, than same-order *AR* model specifications.

Our results have implications for Climate Econometrics and Financial Econometrics models dealing with forecasts at different horizons. As an illustrative example, we show for the Realized Variance of the S&P 500 that while short memory models are well suited for forecasts up to a month ahead, the *ARFIMA* class of models dominate in forecast performance for longer horizons.

## Acknowledgements

The author would like to thank Niels Haldrup for all the insightful comments, the paper greatly improved because of him. The paper was written in part while on a visiting stay at Erasmus University Rotterdam. The author would like to thank the Faculty and Administrative Staff at the Erasmus School of Economics for all their help; in particular to Michel van der Wel and Dick van Dijk for the interesting conversations. This work was supported by CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

## Appendix

### A. Parameters

All	$\varepsilon_t \sim i.i.d.N(0, 1) \forall t$ $T = 1,000; R = 1,000$
<i>DGP 1</i>	$\phi_1 = 0.2$
<i>DGP 2</i>	$N = 10,000; p = 1.4$
<i>DGP 3</i>	$p_k = (\Gamma(k + d)\Gamma(2 - d))/(\Gamma(k + 2 - d)\Gamma(d))$

Table 9: Parameter configuration for the Monte Carlo analysis

## B. Lag Selection Exercise

Model	$d$	ARFIMA	ARMA
DGP 2	0.2	(1,0)	(2,1)
	0.4	(1,0)	(2,1)
DGP 3	0.2	(0,1)	(1,1)
	0.4	(0,1)	(1,2)

Table 10: Results from the lag selection for the *AR* and *MA* components. We show the preferred model for each criteria from  $R = 1,000$  replications using a sample size of  $T = 1,000$ . Results based on the Bayesian Information Criteria following the work of Beran et al. (1998) on consistency of information criteria for long memory processes.

## References

- Andersen, T. G., Bollerslev, T., Diebold, F. X., 2007. Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *Review of Economics and Statistics* 89, 701–720.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71 (2), 579–625.
- Baillie, R. T., 1996. Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73 (1), 5–59.
- Baillie, R. T., Kongcharoen, C., Kapetanios, G., 2012. Prediction from ARFIMA models: Comparisons between MLE and semiparametric estimation procedures. *International Journal of Forecasting* 28 (1), 46–53.
- Barkoulas, J. T., Baum, C. F., 1997. Fractional Differencing Modeling and Forecasting of Eurocurrency Deposit Rates. *The Journal of Financial Research* XX (3), 355–372.
- Beran, J., 1994. *Statistics for long-memory processes*. Chapman & Hall.
- Beran, J., Bhansali, R. J., Ocker, D., 1998. On unified model selection for stationary and non stationary short and long memory autoregressive processes. *Biometrika* 85 (4), 921–934.
- Bhardwaj, G., Swanson, N. R., 2006. An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. *Journal of Econometrics* 131 (1-2), 539–578.
- Chiriac, R., Voev, V., 2011. Modelling and Forecasting Multivariate Realized Volatility. *Journal of Applied Econometrics* 26 (6), 922–947.
- Corsi, F., 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7 (2), 174–196.
- Crato, N., Ray, B. K., 1996. Model selection and forecasting for long-range dependent processes. *Journal of Forecasting* 15 (2), 107–125.
- Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4 (4), 221–238.
- Granger, C. W., 1980. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14 (2), 227–238.
- Granger, C. W., Joyeux, R., 1980. An Introduction to Long Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis* 1 (1), 15–29.
- Haldrup, N., Vera-Valdés, J. E., 2017. Long memory, fractional integration, and cross-sectional aggregation. *Journal of Econometrics* 199 (1), 1–11.
- Hansen, P. R., Lunde, A., Nason, J. M., 2003. Choosing the Best Volatility Models: The Model Confidence Set Approach. *Oxford Bulletin of Economics and Statistics* 65 (401), 839–861.
- Hansen, P. R., Lunde, A., Nason, J. M., 2011. The Model Confidence Set. *Econometrica* 79 (2), 453–497.
- Hosking, J. R. M., 1981. Fractional differencing. *Biometrika* 68 (1), 165–176.
- Hurst, H. E., 1956. The Problem of Long-Term Storage in Reservoirs. *International Association of Scientific Hydrology. Bulletin* 1 (3), 13–27.
- Künsch, H., 1986. Discrimination between monotonic trends and long range dependence. *Journal of Applied Probability* 23 (4), 1025–1030.
- Linden, M., 1999. Time series properties of aggregated AR(1) processes with uniformly distributed coefficients. *Economics Letters* 64 (1), 31–36.



- Martens, M., van Dijk, D., de Pooter, M., 2009. Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. *International Journal of Forecasting* 25 (2), 282–303.
- Oppenheim, G., Viano, M. C., 2004. Aggregation of random parameters ornstein-uhlenbeck or ar processes: Some convergence results. *Journal of Time Series Analysis* 25 (3), 335–350.
- Parke, W., 1999. What is fractional integration? *Review of Economics and Statistics* 81 (4), 632–638.
- Ray, B. K., 1993. Modeling Long Memory Processes for Optimal Long Range Prediction. *Journal of Time Series Analysis* 14 (5), 511–525.
- Robinson, P. M., 1978. Statistical Inference for a Random Coefficient Autoregressive Model. *Scandinavian Journal of Statistics* 5 (3), 163–168.
- Robinson, P. M., 1995. Gaussian Semiparametric Estimation of Long Range Dependence. *The Annals of Statistics* 23 (5), 1630–1661.
- Sowell, F., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics* 53 (1-3), 165–188.
- Zaffaroni, P., 2004. Contemporaneous aggregation of linear dynamic models in large economies. *Journal of Econometrics* 120 (1), 75–102.